

Univerzita Karlova v Praze

Přírodovědecká fakulta

Studijní program: Biologie

Studijní obor: Biologie



Lucie Gillová

Fylogenetické analýzy kulturních znaků

Phylogenetic assessment of cultural characters

Bakalářská práce

Vedoucí bakalářské práce: Doc. RNDr. Daniel Frynta, PhD.

Praha, 2011

Prohlášení

Prohlašuji, že jsem tuto práci vypracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani žádná její podstatná část nebyla použita k získání akademického titulu.

V Praze, dne 17. 8. 2011

Lucie Gillová

Poděkování

Ráda bych poděkovala svému školiči za velkou ochotu a vstřícnost, se kterou mi poskytoval cenné rady a připomínky k mé práci, a za obrovskou trpělivost.

Abstrakt

Fylogenetické metody byly v posledních deseti letech často používány i k analýze kulturních znaků. Tato bakalářská práce nejprve podává stručný přehled a charakteristiku používaných metod a potom se zabývá všemi třemi hlavními oblastmi, kde se dnes používají. První z nich je analýza příbuznosti rukopisů, která je výjimečná tím, že zde dobré výsledky podávají i metody nekladistické. Druhou oblastí, popsanou v práci, jsou data antropologická, zahrnující jak hmotné artefakty, tak různé prvky společenského uspořádání. Poslední oblast se pak zabývá analýzou jazyků a jazykových rodin. Ve všech kapitolách jsou přehledně uvedeny různé přístupy k různým datům a popsány nejvýznamnější studie. Práce se také snaží najít slabiny a klady jednotlivých přístupů.

Klíčová slova: fylogeneze, fylogenetická analýza, lingvistika, antropologie, rukopisy, lexikostatistika, evoluce, kladistika

Abstract

Phylogenetic methods have been used for analysing cultural characters extensively in the last ten years. This bachelor thesis starts with a brief overview and characteristics of the methods used for these analyses and then it deals with all three main fields where these methods have been applied. The first one of them is the analysis of manuscripts and their relations which is unique in that even non-cladistic methods bring usually good results. The second field discussed in this thesis is the anthropological characters, including both artefacts of material culture and various features of social orders. The last one deals with the analysis of languages and language families. In all the chapters, different approaches to different data are listed and the most important studies are described in detail. The thesis also attempts to find negative and positive sides of different approaches.

Keywords: phylogenesis, phylogenetic analysis, linguistics, anthropology, manuscripts, lexicostatistics, evolution, cladistics

Obsah

1 Úvod	2
2 Fylogeneze a fylogenetické metody	3
2.1 Fylogeneze a fylogenetické stromy.....	3
2.2 Fylogenetické a další používané metody	4
3 Fylogenetická analýza rukopisů.....	7
3.1 Paralely evoluce rukopisů a evoluce DNA.....	7
3.2 Kódování dat	8
3.3 Analýzy a výsledky.....	10
3.4 Věrohodnost analýz	11
4 Analýzy antropologických dat	14
4.1 Materiál k analýze	14
4.2 Kódování dat	16
4.3 Analýzy a výsledky	17
5 Fylogenetická analýza lingvistických dat	21
5.1 Evoluce jazyků	21
5.2 Výběr a kódování dat	22
5.3 Metody a analýzy	24
5.3.1 Indoevropská jazyková rodina	24
5.3.2 Austronéska jazyková rodina	26
5.3.3 Jazyky Bantu.....	29
6 Závěr a diskuze	30
7 Použitá literatura	33

1 Úvod

Paralel mezi biologickou a kulturní evolucí si vědci obou disciplín všímají přinejmenším již od publikace Darwinovy knihy *The Origin of Species by Means of Natural Selection* (1859), ve které Darwin poprvé upozornil na fakt, že vývoj jednotlivých jazyků je možné zobrazit stejně jako vývoj jednotlivých druhů, a sice ve formě fylogenetického stromu (Darwin, 1859). Se současným chápáním fylogeneze a způsobem, jak korektně konstruovat strom, ale přišel až v roce 1965 Willi Hennig. Navíc se biologie a lingvistika vyvíjely nezávisle, a ačkoliv zobrazení vztahů pomocí stromů používaly obě dvě, k syntéze poznatků a využití metod vyvinutých primárně pro biologii došlo až na přelomu dvacátého a dvacátého prvního století, kdy byly publikovány první práce, používající ke stanovení vztahů v rámci jazykových rodin počítačové metody, vyvinuté primárně pro biologii, namátkou Gray & Jordan (2000), Gray & Atkinson (2003), Holden (2002), Rexová et al. (2003), a později i mnohé další. Navíc byly fylogenetické metody zhruba ve stejné době poprvé použity i na další kulturní data. Jedná se o dvě hlavní oblasti. První z nich je analýza příbuznosti středověkých a raně novověkých manuskriptů. Tato aplikace se přímo nabízí, vzhledem k podobnosti evoluce rukopisů a evoluce DNA (Howe et al., 2001), což bude demonstrováno dále. Poslední zásadní oblastí je aplikace fylogenetických metod na analýzu příbuznosti u artefaktů hmotné kultury a dalších antropologických dat. Zde byly provedeny například fylogenetické analýzy na základě vzorů perských koberců nebo vzorů na keramice (Tehrani and Collard, 2002; Cochrane and Lipo, 2010). Tato oblast zkoumání není tak rozsáhlá, protože je těžší než ve dvou předchozích vybrat si a shromáždit použitelná data. Navíc se ukazuje, že se zde v mnohem větší míře projevují spoje mezi přímo nepříbuznými větvemi a tedy síťovitá struktura, způsobená například přímým kontaktem mezi nepříbuznými populacemi. Kromě fylogenetických analýz se provádějí i analýzy nefylogenetické, založené na některé z clusterovacích metod, jako jsou UPGMA a Neighbor-Joining.

Tato práce by měla shrnovat nejzásadnější poznatky všech výše zmíněných oblastí a ukázat, jak přínosné mohou primárně biologické metody být i při aplikaci na data kulturní, už proto, že kulturní data nám mohou poskytnout cenné informace o vývoji lidské populace. Zároveň by měla ukázat, jak zásadní je výběr dat, a na základě simulačních studií (například

Barbançon et al., 2006) ukázat i to, že rovněž výběr použité metody je zásadní a každá metoda má své limity.

2 Fylogeneze a fylogenetické metody

2.1 Fylogeneze a fylogenetické stromy

Fylogeneze je definována jako evoluční vývoj druhu nebo jiné skupiny organismů přes jednotlivé vývojové stupně. V tomto smyslu byl termín poprvé použit Ernstem Haeckelem v roce 1866. Vzhledem k analogii mezi evolucí organismů a evolucí kulturní se tento termín dnes používá pro vývoj, ve kterém se jednotlivé linie větví, mění a spojují, také v širším slova smyslu. V obou oblastech totiž dochází ke kladogenezi, tedy větvení jedné větve na více nových, anagenezi, tedy vývoji v rámci taxonu bez větvení, stazigenezi, tedy procesu nulové anageneze, přetrvávání taxonu beze změn, a syngenezi, tedy spojování jednotlivých větví.

Fylogenetika je věda, která se snaží rekonstruovat fylogenezi. Kladistika je metoda pro studium kladogeneze, poprvé ji definoval Willi Hennig a nazval ji fylogenetickou systematikou (Hennig, 1965). Definoval, jak korektně tvořit fylogenetický strom a které znaky jsou pro jeho tvorbu podstatné. Kladistika pracuje s binárním větvením a znaky, které při rekonstrukci používá, jsou pouze apomorfie, což jsou odvozené znaky nevyskytující se u předků. Nikdy nepoužívá plesiomorfie, primitivní znaky přítomné i u předků daného taxonu. Rekonstrukce tedy neprobíhá na základě podobnosti koncových větví. S konkrétním způsobem fylogenetické analýzy a využitím konkrétních dat přišli ale až po více než dvaceti letech Harvey a Pagel (Harvey & Pagel, 1991). V současnosti je klasifikace a systematika organismů založena právě na fylogenezi a kladistickém přístupu. Metody používané pro analýzu kulturních znaků ale nejsou vždy striktně kladistické, naopak, například metoda Neighbour-Joining je v zásadě metodou fenetickou a některé práce používají právě ji. Fenetický přístup se zaměřuje na vnější podobnost a už neřeší, jak v evoluci vznikla. Právě díky nerozlišování apomorfii a plesiomorfii často podává zkreslené výsledky, například pokud v rámci analyzovaných dat došlo ke konvergentní evoluci.

Fylogenetika pro zobrazení vztahů mezi organismy používá fylogenetické stromy. Zobrazení pomocí stromů bylo používáno už dříve, ovšem data a metody používané pro rekonstrukci byly jiné (Atkinson and Gray, 2005). Fylogenetický strom se skládá z postupně

se odvětvujících větví, které spojují jednotlivé uzly. Koncové uzly představují konkrétní taxony, v rámci kulturních dat například jednotlivé jazyky. Větvení v rámci stromu je obvykle binární, to znamená, že jeden společný předek dává vznik dvěma dceřinným taxonům. Pokud větvení binární není, znamená to, že je strom nevyřešený (unresolved) a k plnému řešení chybí data nebo není analýza zcela správná. Toto ovšem u kulturních dat nemusí platit vždy. Strom může být zakořeněný nebo nezakořeněný. Zakořeněnost znamená, že známe polohu společného předka všech zahrnutých taxonů. V praxi se strom zakořeňuje nejčastěji přidáním takzvané vnější skupiny (outgroup), o které víme, že je příbuzná se společným předkem, ale do námi analyzované skupiny taxonů nepatří. V analýzách kulturních dat se ale často objevují i stromy nezakořeněné, právě kvůli obtížnému určení vnější skupiny nebo díky tomu, že vnější skupina ani neexistuje (to je případ některých rukopisů, pokud nemáme k dispozici autorskou verzi, můžeme se jen dohadovat, kde by měl být kořen- Barbrook et al. 1998).

Druhým možným zobrazením jsou sítě. Sítě mají tu výhodu, že umožňují zobrazit i jiné vztahy, například výpůjčky v jazyce nebo horizontální transfer v biologii. Sítí je opět několik typů. Explicitní fylogenetické sítě zobrazují jasný evoluční scénář. Dá se v nich najít základní fylogenetický strom, stejně jako kontaktní události v evoluci. Druhým typem je síť, která je založená na zobrazení rozdělení (split) a nezobrazuje konkrétní evoluční scénář, spíše jde o zobrazení kombinace stromu a signálů, které strom ruší (Nichols and Warnow, 2008). Navíc v tomto typu grafu vnitřní uzly neoznačují společné předky, ale spíše konflikt mezi možnými různými rozděleními. Interpretace takové sítě pak může být složitá, protože často nevíme, jestli jde o kontaktní události nebo pouze o nedostatečná nebo špatně zvolená data. Na druhou stranu byly sítě při analýze kulturních dat použity mnohokrát a ukázalo se, že při řešení určitých problémů mohou být přínosné (Bakker et al., 2011, Nakleh et al., 2005 atd.).

2.2 Fylogenetické a další používané metody

K analýze dat a získání stromu bylo pro kulturní data použito několik metod. Nejpoužívanější jsou metody Maximum Parsimony (MP), Maximum Compatibility (MC), Maximum Likelihood (ML) a Bayesiánská analýza, které patří mezi metody založené přímo na diskrétních datech, což znamená, že jednotlivé znaky v analýze zůstávají zachovány. Dále se používají i metody založené na vzdálenosti mezi jednotlivými taxony, mezi které patří UPGMA (unweighted pair group method with arithmetic mean) a Neighbor-Joining. Striktně kladistická je z těchto metod pouze metoda maximální parsimonie.

UPGMA (Sokal & Michener, 1958) je metoda používaná v lexikostatistice, v současnosti už se pro analýzu jazykových dat příliš nepoužívá, především proto, že je nefylogenetická, původně byla vyvinuta pro konstrukce fenogramů. V každém kroku spojuje dva nejbližší taxony a postupně tak konstruuje strom, jedná se tedy o metodu clusterovací. Vychází ale z předpokladu, že se evoluční procesy řídí lexikálními/molekulárními hodinami a evoluce tedy u všech taxonů probíhá stejným, neměním se tempem. To však neplatí, a proto jsou výsledky získané touto metodou značně nepřesné (Barbançon et al., 2006).

Neighbor-joining (Saitou & Nei, 1987) je druhá metoda založená na vzdálenostech mezi taxony. Je také používána podstatně častěji, protože v rámci této skupiny bývá považována za jednu z nejlepších (Barbançon et al., 2006). Je to však stále jenom analýza clusterová (Wardova metoda) a pravděpodobně nerekonstruuje skutečnou fylogenezi. Na rozdíl od UPGMA nespojuje nejméně vzdálené taxony, místo toho počítá přeměnu dané distance matrix a poté spojuje dva taxony, mezi kterými je tato hodnota nejnižší. Pak už pracuje podobně, spojí dva taxony, které po spojení bere jako taxon jeden a hledá k němu jiný nejbližší.

Maximum Parsimony, metoda maximální parsimonie, je jednou z nejpoužívanějších metod při analýze kulturních dat. Hledá takový fylogenetický strom, na kterém se v rámci vývoje objevilo nejméně evolučních změn, hledá tedy cestu nejjednodušší a nejkratší. Problémem při použití této metody je hlavně to, že je výpočetně velmi náročná, takže se u ní, stejně jako u metod následujících, používá heuristika, aby bylo nalezení vhodného stromu vůbec možné. Pro tuto metodu se používá nejčastěji program PAUP* (Swofford, 1996). Je to jediná striktně kladistická metoda.

Maximum Compatibility je metoda hledající strom, na kterém se co nejvíce znaků vyvíjí bez homoplázií. Stejně jako u metody maximální parsimonie, i u této metody je možné dodat jednotlivým znakům váhu (weighted MP nebo MC), když víme, že se jednotlivé znaky vyvíjely různě a u některých skupin se dá předpokládat, že k homoplázii docházelo v podstatně menší míře. Toto řešení ale odporuje striktní kladistice, která předpokládá, že v evoluci všechny znaky vznikají se stejnou pravděpodobností a pouze jednou. K homopláziím tedy nedochází a ve chvíli, kdy stejný znak vznikne víckrát, už to není ten samý znak. Proto je i nevážená MC v zásadě nekladistická. Obě vážené metody se přesto používají. Barbançon et al. (2006) takto například odlišili data morfologická, s váhou 50, a lexikální s váhou 1. MC stejně jako MP často najde stromů několik, se stejnou hodnotou,

takže se pak výsledek vyjadřuje jako takzvaný consensus tree, tedy strom, ve kterém jsou větvení, která se nachází na většině výsledných stromů.

Maximum Likelihood je metoda, která je, stejně jako Bayesiánská analýza, založená na explicitních parametrických modelech evoluce. Jednou z výhod těchto přístupů je fakt, že předpoklady o tom, jak evoluční proces v daném případě funguje, jsou na začátku vyslovené. ML se snaží nalézt strom a s ním spojené parametry tak, aby maximalizovalo pravděpodobnost, že vzniknou právě pozorovaná data (Nichols and Warnow, 2008). Díky parametrům tato metoda někdy překonává MP, ale v případě jejich špatného zvolení vytváří chybné stromy. Proto není vhodné ji používat v případech, kde parametry není možné správně určit.

Bayesiánská analýza je vlastně založena na ML, ale je komplikovanější. Začíná stejně jako ML s určitými parametry, ale zároveň i s určitým modelovým stromem, od něhož pak pokračuje náhodnou cestou a vytváří stromy další. Pokud je nový strom pravděpodobnější než strom původní, dojde k záměně, pokud ne, zůstane jako lepší možnost strom původní. Analýza se postupně dostane do stacionární fáze, kdy už se pravděpodobnost nijak výrazně nezlepšuje, a strom můžeme považovat za výsledný. K úplně ideálnímu stromu se ale vzhledem k výpočetní náročnosti metody není možné dostat.

Pro určení přesnosti analýzy a věrohodnosti výsledků se používá několik metod. Bootstrapping je statistická metoda, která vytváří náhodné datové soubory z původního souboru (stejně velké) a potom pro každý takový soubor konstruuje strom. Následně se určuje frekvence, s jakou se každá jednotlivá větev stromu původního vyskytne i v náhodných datech a udává se v procentech nebo jako číslo 0-1. Čím vyšší hodnota, tím větší podpora, ale vzhledem k rozdílnosti jednotlivých analýz stále nepanuje shoda, jak vysoká musí hodnota být, aby daná větev měla podporu. Kromě bootstrappingu se používá ještě index konzistence, který je vlastně měřítkem počtu homoplazií. Je to poměr minimálního počtu změn nutných k vytvoření kladogramu a počtu změn skutečně pozorovaných na konkrétním kladogramu. Pokud je index 1, data přesně souhlasí s kladogramem. Opět tedy platí, že čím vyšší index, tím spíše je větev správně. Index retence je poměr synapomorfí v datech a synapomorfí, které jsou na kladogramu skutečně zachovány. Platí, že vyšší hodnota znamená vyšší podporu. U obou indexů ale není přesně stanovaná hranice, nad kterou by výsledky měly být správné.

3 **Fylogenetická analýza rukopisů**

3.1 **Paralely evoluce rukopisů a evoluce DNA**

Fylogenetické analýzy rukopisů podávají dobré výsledky, což bylo prokázáno i analýzami uměle vytvořených dat (Spencer et al., 2004; Baret et al., 2006). Jak rukopisy, tak jednotlivé druhy pocházejí z jednoho společného předka, takže i mezi evolucí rukopisů a evolucí DNA lze najít mnoho paralel (Howe et al., 2001). Ve středověku a raném novověku, před vynálezem knihtisku, bylo možné vytvářet více kopií textu pouze přepisem. Písaři ale nebyli neomylní a navíc občas v textu provedli i záměrné změny, v místech, která jim připadala chybná, nebo naopak v místech, kde použili variantu výrazu typickou pro jejich region. Písař text vždy opisoval podle určité předlohy a fylogenetickou analýzou je tedy možné zjistit příbuznost jednotlivých verzí textu. Už před použitím fylogenetických metod samozřejmě vědci analyzovali příbuznost textů, dělali to ale přímo srovnáváním jednotlivých textů, což je pracnější a u rozsáhlejších textů špatně proveditelné. Ačkoliv se i v tomto odvětví objevovaly snahy vytvořit vlastní počítačové programy, použití programů určených primárně pro biologii se nabízelo. Paralely, které ukazují proč, ve své práci shrnují Howe et al. (2001). První studie z oblasti fylogeneze rukopisů se navíc objevují už před aplikací těchto metod v lingvistice a archeologii (Platnick & Cameron, 1977; Lee, 1989; Robinson & O'Hara, 1996), ačkoliv ještě nepoužívají stejné počítačové metody jako práce současné.

Asi nejvýraznější paralelou mezi evolucí DNA a evolucí rukopisů jsou bodové mutace, v textech se nejčastěji mění jeden výraz, ať už je to způsobeno subjektivním výběrem písaře nebo chybou a přehlédnutím se při přepisu. Existují ale i další paralely, které se v evoluci rukopisů projevují, i když v menší míře. První z nich je rekombinace, u rukopisů to znamená, že písař během přepisu z nějakého důvodu změnil předlohu (exemplum) (Howe et al., 2001). Výsledkem je text, který vznikl přepisem dvou různých předloh a který není vždy snadné ve fylogenetické analýze správně zařadit. Kromě rekombinace může v evoluci rukopisů docházet i k laterálnímu transferu, a to tehdy, když kromě samotného textu existuje ještě jiný, podobný. Pak může písař úmyslně, například pro dovysvětlení, přidat přepisovaného textu část z toho druhého. Howe et al. (2001) uvádějí jako příklad text *Kings of England*, jehož autorem je Lydgate. Kromě tohoto textu ještě existuje *Kings of England II*, který se liší. Písař při přepisu Lydgateova textu převzal i několik řádek z textu druhého. Podobně přepisovali písaři i text *Treatise on the Astrolabe* od Chaucera, který je sám o sobě kompilátem textů starších a navíc

je nedokončený (Eagleton & Spencer, 2006). Dále se objevuje i delece, až celých řádků, ke které dochází pravděpodobně písařovým přehlédnutím. Stává se to, když dva řádky končí stejným slovem a písař místo na první již přepsaný naváže na druhý (Howe et al., 2001). Může také docházet ke konvergentní evoluci- použití nářeční varianty apod. Tak se ve dvou textech mohlo například objevit slovo kirk místo church, ačkoliv jinak tyto verze spojené nejsou (Howe et al., 2001). Kromě těchto základních paralel na úrovni slov je někdy nutné pracovat i na úrovni vyšší. To je případ některých rukopisů, které se skládají ze samostatných kapitol. Řazení těchto kapitol může být velmi proměnlivé, což je případ *The Canterbury Tales* (Spencer et al., 2003). V takových případech je možné analyzovat příbuznost na základě pořadí kapitol.

3.2 Kódování dat

Díky těmto paralelám by měly fylogenetické metody fungovat dobře. Na druhou stranu, jak uvádějí Howe et al. (2001), je někdy problém už se samotnou dostupností materiálu, který je nutné mít v digitální podobě. To není vždy snadné, kvůli dostupnosti jednotlivých rukopisů i kvůli náročnosti přepisu. Pokud už jsou dostupná data v digitalizované podobě, kódování do podoby, která by se dala v programu analyzovat, probíhá různě. Závisí to na délce textu, který chceme analyzovat, na předpokládané evoluci, na tom, jak velký text chceme analyzovat, jaká je předpokládaná evoluce a jaký program je tedy dobré použít. Ve studii Howe et al. (2001) byl použit písmenný kód pro jednotlivé změny. Každý řádek v takovém kódování představuje jinou verzi textu a každý sloupec určitou pozici v textu. Pokud se texty neliší, jsou v případě Howe et al. (2001) značeny písmenem A, pokud je řádek změněn zcela, jsou to písmena B, C atd., pokud je část řádku vynechána, je to písmeno F apod. Někdy se kóduje pouze rozdílnost. Binární kódování použili Robinson a O'Hara (1996), pro každé místo zaznamenávali přítomnost 1/ nepřítomnost 0 znaku. Nicméně pouze dvě varianty znaky jsou u rukopisů vzácné a binární kódování se už příliš nepoužívá, například ve studii Lantin et al. (2004) se hodnoty pohybují od 0, což znamená žádnou změnu, až po 7. Toto kódování, ať už číslíkové nebo písmenné, je asi nejčastější, protože takto zpracovaná data jsou použitelná pro většinu programů. Inzerce a delece se nejčastěji kódují jako změny, tedy pouze jiným symbolem (Windram et al., 2008; Eagleton & Spencer, 2006). Záznam více po sobě jdoucích

vynechaných slov je složitější. Pokud každé slovo budeme brát jako samostatnou událost a pokud tím pádem budeme každé slovo kódovat zvlášť, může prosté vynechání řádku dostat přehnaný význam a ovlivnit analýzu (Eagleton & Spencer, 2006). Windram et al. (2008) oba možné způsoby kódování inzercí a delecí porovnali a výsledky se příliš nelišily. Ale toto zacházení s inzercemi a delecemi se týká jednotlivých slov, v případě analýzy podle pořadí jednotlivých částí textu se Spencer et al. (2003) rozhodli vynechané části rukopisu v analýze nepoužít ani tam, kde se vyskytly. Nebylo totiž možné určit, které části v textu původně opravdu nebyly a které se jen nedochovaly. Toto může být někdy i případ větších delecí v rámci jednotlivých řádek textu, takže části, kde byl některý rukopis prokazatelně zničen, by měly být z analýzy vyjmuty.

Dále záleží na tom, jakou metodu je vhodné použít. Při použití metod, které jsou založené na diskrétních datech, lze pracovat už přímo s výše uvedenými kódovanými datovými soubory (Lantin et al., 2004). Windram et al. (2008) použili metodu maximální parsimonie i Neighbor-Net, druhá jmenovaná je založená na vzdálenosti, ale výsledky se příliš nelišily. Většina studií v této oblasti používá právě metody založené na vzdálenosti mezi jednotlivými páry v rámci souboru dat, ne přímo na znacích samotných. Spencer et al. (2003) argumentují tím, že pro metody založené přímo na znacích by byl objem dat příliš velký a čas potřebný pro analýzu příliš dlouhý. Ačkoliv jde tedy o metody nefylogenetické, v tomto případě je pravděpodobné, že skutečně budou rekonstruovat fylogenezi, protože při opisování se horizontální transfer ani konvergentní evoluce nemohou ve větší míře uplatnit.

Použití těchto metod s sebou ale přináší nutnost vzdálenost mezi rukopisy nějak spočítat. Nejčastěji se počítá jako podíl změn mezi dvěma rukopisy a celkového počtu znaků, což znamená, že jde o číslo pohybující se od 0 do 1 (např. Eagleton & Spencer, 2006). Spencer & Howe (2001) ve své práci přišli s alternativním způsobem výpočtu vzdáleností. Vycházejí z toho, že určité změny, které pozorujeme, mohou být jen začátek a konec postupného vývoje, ale my nemáme možnost tyto mezistupně vidět. Aby byla analýza přesnější, je potřeba do ní nějak započítat právě i mezistupně. Spencer & Howe (2001) vyšli z Jukes-Cantorova modelu a přizpůsobili ho pro potřeby analýzy rukopisů. Při porovnání výsledků za použití neopravených vzdáleností a asymptotických Juke-Cantorových (AJC) vzdáleností se ukázalo, že model zůstal téměř stejný, nejpodstatnější rozdíl byla délka jednotlivých větví, která se u AJC vzdáleností značně prodloužila. Spencer a Howe ale použili pouze jeden řádek básně, pro další závěry jejich práce tedy rozhodně nestačí. Tuto metodu použili ještě Spencer et al. (2003), ale už zde není srovnání s neopravenými vzdálenostmi,

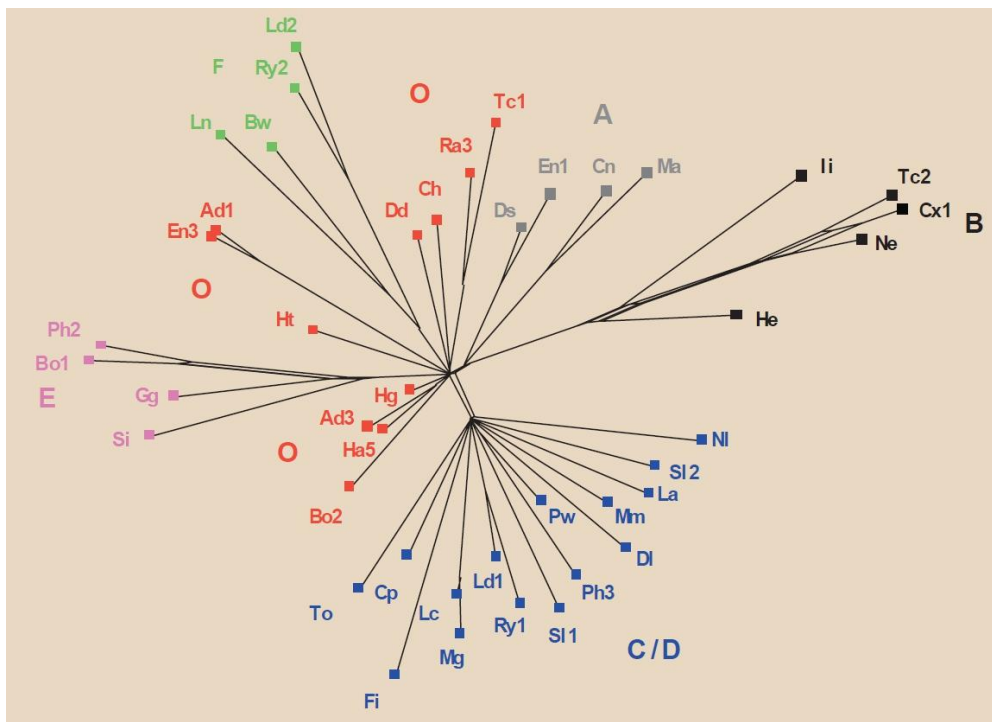
navíc práce se zaměřuje na minimální počet znaků, který je nutný pro spolehlivou analýzu. Pro analýzu z pořadí jednotlivých *Canterburských povídek* Spencer et al. (2003) použili jinou metodu, předchozí výpočet vzdáleností by se pro jejich data nehodil. Jejich vzdálenosti jsou předpokládané počty změn nutných pro přeměnu jednoho rukopisu v jiný. Použili k tomu metodu vzdálenosti bodu zlomu (breakpoint distance), což je počet znaků, jejichž soused napravo je jiný.

3.3 Analýzy a výsledky

Jednotlivé soubory dat jsou potom analyzovány příslušnými programy, podle toho, jaká metoda se na daný typ dat nejlépe hodí. Poměrně často jsou používány programy SplitsTree (Huson, 1998), Neighbor-Net (Bryant & Moulton, 2004) a PAUP* (Swofford, 1996). Některé studie (Barbrook et al. 1998, Windram et al., 2008) používají pro srovnání programů více. Výsledky pak ukazují jednak příbuznost jednotlivých rukopisů a jednak to, jak blízko původní verzi jednotlivé rukopisy jsou, alespoň tehdy, kdy je délka větví úměrná počtu změn. Barbrook et al. (1998) takto analyzovali ‘The Wife of Bath’s Prologue’ z Chaucerových *Canterburských povídek*. Podařilo se jim dokázat, které texty jsou nejbližší původnímu, nedochovanému, textu (obr.1). Zároveň také obrovské množství změn i v tak krátkém textu znovu potvrdilo, že v případě *Canterburských povídek* písaři nespíš vycházeli z nedokončeného autorova rukopisu, kde mohlo být množství poznámek k tomu, co vynechat, přidat, nahradit a podobně, takže výsledné verze byly spíše výběrem písaře.

Canterburské povídky jsou pro analýzy velmi vhodný materiál, rukopisy jsou totiž dobře dostupné v digitalizované podobě. Navíc jsou tyto analýzy zajímavé, protože u díla neznáme ani zamýšlené pořadí jednotlivých povídek a můžeme si většinou jen domýšlet, jak je Chaucer plánoval propojit. Analýzou příbuznosti na základě pořadí jednotlivých povídek se zabývali Spencer et al. (2003). Protože ne všechny rukopisy obsahují všechny povídky, byly do výsledného souboru zakódovány jen povídky, které jsou všem rukopisům společné. Po srovnání výsledku s předpokládaným pořadím rekonstruovaným na základě manuálního porovnání došli Spencer et al.(2003) k podobným závěrům, fylogenetická analýza odhalila většinu předpokládaných skupin. Navíc po srovnání s analýzou založenou na jiných datech, konkrétně na prvních 500 řádcích z ‘Prologu,’ opět došlo k potvrzení základního větvení. *Canterburské povídky* jsou používané nejčastěji, analyzuje se fylogeneze na základě

jednotlivých povídek, na základě pořadí povídek. Byly použity i další biologické metody, pro určení změny vzoru (Windram et al., 2005).



Obrázek 1 SplitsTree- analýza 44 rukopisů 'Wife of Bath's Prologue', Barbrook et al. (1998)

Poměrně specifickou analýzou je práce Skeltona (2008), který používal zcela jiná data než ostatní studie. Rekonstruoval fylogenezi podle stylu psaní jednotlivých písmen rukopisu, protože pracoval s lineárním B písmem (obr. 2). Výsledky jsou konzistentní s předchozími rozbory a zároveň potvrzují některé problematické otázky, takže i na takovýto typ dat se fylogenetické metody dají aplikovat, ačkoliv už jsou to data možná bližší analýzám hmotné kultury, třeba vzorům na keramice, než klasické analýze rukopisů.

3.4 Věrohodnost analýz






















U většiny studií je možné srovnání s předchozími pracemi, které byly provedeny klasicky. Někde navíc u příbuznosti panuje mezi vědci vcelku shoda, například u norských ság (Robinson & O'Hara, 1996). I tak je ale dobré ověřit si, jak fylogenetické metody vlastně fungují, čímž se zabývalo několik studií. Spencer et al.(2004) ve své práci použili 8 odstavců

z německé básně Parzival od Wolframa von Eschenbacha, které nechali postupně opisovat 20

Table Describing Characters and Character States

	Char. 46	Char. 47	Char. 48	Char. 49	Char. 50	Char. 51	Char. 52	Char. 53	Char. 54	Char. 55
Hagia Triada Linear A	0	1	1	1	0	0	-	1	-	-
RCT 124-B	0	0	2	2	2	1	1	0	2	1
Knossos 101	2	0	2	0	1	1	0	0	2	1
KN V52	2	0	2	0	?	?	?	?	?	?
Mycenae 51	0	0	?	?	0	1	0	0	0	2
Pylos 1	0	0	2	0	0	1	0	2	2	2
Thebes 303	0	0	0	0	0	1	0	0	2	0

Table Describing Characters and Character States

46	AB30 NI a	 0	No circle where Y branches	1	Scribe uses both	 2	Circle where Y branches
47	AB30 NI b	 0	One stroke crossing each arm of the Y	 1	Two strokes cross each arm of the Y		
48	AB30 NI c	 0	Two branches meet at the bottom of the sign (ie, there is no stalk)	1	Scribe uses both	 2	Two branches meet midway up the stalk
49	AB30 NI d	 0	Y drawn in two strokes	1	Scribe uses both	 2	Y drawn in three strokes
50	AB31 SA a	 0	No circle where Y branches	1	Scribe uses both	 2	Circle where Y branches
51	AB31 SA b	 0	No side ticks	 1	Has side ticks		
52	AB31 SA c	 0	Without extra side ticks	 1	With extra side ticks		
53	AB31 SA d	 0	Y drawn in two strokes	1	Scribe uses both	 2	Y drawn in three strokes
54	B32 QO a	 0	No crossbar	1	Scribe uses both	 2	Crossbar
55	B32 QO b	 0	One continuous straight line	 1	One continuous wavy line	 2	Two separate lines
		3	Scribe uses both 0 and 1	4	Scribe uses both 0 and 2	5	Scribe uses both 1 and 2

Obrázek 2 Skelton (2008)- ukázka datového souboru spolu s jednotlivými znaky

studentů. Náhodně při tom vybírali verzi (původní nebo některou z již přepsaných), kterou dávali jako předlohu. Studenti o typu experimentu informováni nebyli. Tímto způsobem Spencer et al. (2004) získali soubor rukopisů s předem známým vývojem a mohli tak na něm otestovat, jak metody fungují. Testovali takto metodu maximální parsimonie, Neighbor-Joining, split decomposition, Neighbor-Net i výpočet vzdálenosti, který navrhli Spencer & Howe (2001). Výsledek nejbližší skutečný fylogenezi pak byl vytvořen pomocí metody Neighbor-Joining, i když maximální parsimonie se od něho příliš nelišila. To naznačuje, že u rukopisů by k rekonstrukci skutečně mohla stačit i fenetická analýza, protože podobnost značí příbuznost a horizontální transfer a konvergentní evoluce hrají roli minimální. Dvě nejuspěšnější metody měly problém hlavně s rekonstrukcí vnitřních uzlů, protože obě předpokládají, že rukopisy jsou koncové body. Na druhou stranu ve všech případech měly nesprávně přiřazené větve nízkou podporu (bootstrap support nižší než 70%). Z toho se dá usuzovat, že v podobných případech je často nutná bližší analýza a nedá se plně spoléhat na výsledky těchto dvou metod. Metoda split decomposition dokázala rozlišit jen některé skupiny. Metoda Neighbor-Net poskytla naopak výsledky dobré, je ale těžké interpretovat, co jednotlivá propojení větví znamenají. Bayesiánská metoda a metoda Maximum Likelihood zatím použity nebyly, kvůli nutnosti určit parametry. Baret et al. (2006) postupovali podobně a došli ke stejným závěrům. Upozorňují opět na fakt, že fylogenetické metody nejlépe fungují v kombinaci s klasickou analýzou, jednak kvůli internalizaci některých uzlů, jednak kvůli identifikaci případných polytomií. Metody totiž pracují s binárním větvením, ale jeden rukopis mohl být a nejspíš byl kopírován více než dvakrát. Počtem znaků, potřebným pro spolehlivou analýzu, se zabývali Spencer et al. (2003), ale i nejmenší použitý počet – 123 znaků - stačil pro rekonstrukci fylogeneze, byť ne zcela přesné. V této práci je ale diskutabilní použití textu s ne zcela známou fylogenezí. Výsledky pak v tomto případě nemohly být porovnány se skutečnou fylogenezí, ale s dendrogramem vytvořeným pomocí největšího počtu znaků. A logické je, že čím víc znaků, tím blíže „vzorovému“ dendrogramu.

Další problematickou částí je výše zmiňovaná „rekombinace“, změna vzoru, který písař opisoval. V samotné analýze se toto dá podchytit rozdělením datového souboru na několik částí a porovnáním výsledku pro jednotlivé části (Windram et al., 2008; Phillips- Rodriguez et al., 2009). Pokud jsou v některé části rozdíly v umístění rukopisů, je možné, že došlo ke změně vzoru. Pro určení, kde ke změně vzoru došlo, se osvědčila matematická metoda- chí-kvadrát test. Používá se obecně ke zjištění, zda data odpovídají pravidelnému rozdělení. V případě rukopisů pak program předpokládá, že od určitého bodu by mělo být na obě strany

určité množství změn (dle polohy bodu). Pokud se toto rozdělení výrazně liší, znamená to, že došlo pravděpodobně ke změně předlohy a v místě, kde je nejvyšší hodnota chí-kvadrátu, je nejpravděpodobněji bod zlomu. Autoři tuto metodu obvykle srovnávají s přímou analýzou textů (Phillips- Rodriguez et al., 2009).

Fylogenetické analýzy rukopisů jsou tedy nejstarší oblastí kulturních dat, na kterou byly tyto metody aplikovány. Několik studií, které testují jejich účinnost, prokázalo, že metody podávají velice dobré výsledky, i když je nutné uvědomit si určitá omezení. V současnosti je tedy stále potřeba přistupovat k výsledkům kriticky a případně je doplnit klasickou stematologickou analýzou.

4 Analýzy antropologických dat

4.1 Materiál k analýze

I v antropologii se fylogenetické analýzy používají, i když v podstatně menší míře než v lingvistice, vzhledem k obtížnosti výběru analyzovatelného materiálu. Rozptýl dosud analyzovaných věcí je obrovský, v podstatě od jednotlivých sociálních systémů a jejich vývoje po hmotné artefakty lidské kultury jako jsou keramika, textilie, ale i hroty střel. Problematické je totiž určení znaků, které by pomohly odhalit skutečnou fylogenezi, a rovněž často spíše síťovitá výsledná struktura, která naznačuje, že některé kulturní prvky se šířily spíše kontaktem mezi jednotlivými populacemi, tj. horizontálním transferem (např. Cochrane & Lipo, 2010). Navíc se v antropologii tyto metody nepoužívají pouze k rekonstrukci fylogeneze, antropology spíše zajímá možnost ověřit si souvislost některých prvků kultury. Borgerhoff-Mulder et al. (2001) tak například zkoumali souvislost různých kulturních prvků souvisejících s manželstvím (například velikost věna/toho, co muž musí zaplatit za ženu, možnost rozvodu apod.) s dalšími faktory, které by je mohly ovlivňovat (velikost stáda, polygynie atd.).

Do současnosti byly analyzovány nejrozumnější předměty a data. Jedna z prvních analýz se zabývala výše zmiňovanými prvky, které souvisí se sociální strukturou společností. Na druhou stranu v porovnání s klasickou metodou nebyla tak úspěšná. Dalším z analyzovaných předmětů jsou hroty střel, kterými se zabývá hned několik prací (O'Brien et al., 2001, 2010, Buchanan & Collard, 2007). Na nich se určují především prvky kvantitativní, rozměry

jednotlivých částí, a potom také tvar. Problémem těchto analýz je ale nejistota, kolik používaných znaků je skutečně nezávislých. Pokud autoři pracují s rozměry hrotů, je pravděpodobné, že přinejmenším některé z nich spolu budou korelovat a pak se už tak malá sada znaků zmenšuje tak, že výsledky analýzy se dají dost zpochybnit. Dále se dá pracovat s až téměř uměleckými prvky. Vzor na keramice je jedním z nich, Cochrane & Lipo (2010) takto analyzovali keramiku z Oceánie. Používali jednotlivé identifikované vzory (rovné čáry, přerušované čáry, jejich jednotlivé varianty a podobně). V této analýze bylo použito 107 různých vzorů, takže i kdyby některé z nich byly korelované, soubor dat je pravděpodobně stále dostatečně velký. Vzory na tkaných textiliích se v několika pracích zabýval Tehrani a Collard (Tehrani & Collard, 2002- turkmenské textilie, Tehrani & Collard, 2009, Tehrani et al. 2010, Matthews et al., 2011- íránské textilie)- obr.3. Stejně jako u keramiky, i zde byly identifikovány základní prvky vzorů. Shennan & Collard (2005) dokonce pracovali se souborem nejrůznějších artefaktů (náušnice, hřebeny, koše, štíty apod.) a vycházeli z předpokladu, že přítomnost těchto předmětů značí přítomnost určitých společenských zvyklostí. Těmkin & Eldredge (2007) analyzovali dva typy hudebních nástrojů, a jejich analýza ukazuje právě nutnost určit si správná data. Posledním odlišným typem dat byly záznamy o výskytu jednotlivých hospodářských rostlin (Coward et al., 2008).



Obrázek 3 Matthews et al. (2011)- konkrétní varianty používané v analýze

4.2 Kódování dat

Kódování v případě antropologických dat bývá podobné jako u rukopisů, kóduje se přítomnost nebo nepřítomnost určitého prvku. Jde tedy často o kódování binární. Tehrani a Collard (2002) pracovali se vzorkem 60 tkaných textilií, na kterých si určili 90 znaků a u těch pak kodovali jejich přítomnost, respektive nepřítomnost. I Coward et al. (2008) vycházejí z dat, kde byla odlišena i míra výskytu jednotlivých hospodářských plodin (stupnice 1-4), ale do analýzy zahrnují jen vysoký výskyt plodin, tedy 1, nebo nepřítomnost 0. U těch znaků, které nemají jen dva základní stavy, binární kódování nestačí. Ve studiích, které pracovaly s kvantitativními daty, tedy bylo potřeba zakódovat stavů více. To se týká například práce O'Brien et al. (2001), kteří pracovali s hroty střel, na kterých určovali celkem 8 odlišných znaků, z nichž pouze u jednoho byla kódována jen přítomnost/nepřítomnost. U dalších znaků to bylo kódování tvaru špičky a základny i přímo míry, jednak úhly (stanovená rozmezí) a jednak poměr šířky a délky hrotu- opět jednotlivá čísla pro rozmezí, například 1-1,99 bylo kódováno jako 1, 2-2,99 jako 2 atd. Největší počet variant jednoho znaku byl 6. Stejně jako u rukopisů, i u antropologických dat pomocí takového kódování vznikne soubor, kde řádky jsou jednotlivé předměty a sloupce jednotlivé hodnoty v různých znacích. I Borgerhoff-Mulder et al. (2001) ve své práci kódují některé znaky jako jednotlivé stavy (od 0 do 5 nejvíce). Zároveň ale u některých znaků uvádějí přímo jejich hodnoty- například cena, kterou muž platí za manželku, je sice převáděna na jednu jednotku (tj. započítává se rozdíl mezi hodnotami jednotlivých zvířat), ale udává se přímo číselně, není zařazována do jednotlivých očíslovaných kategorií, jako je tomu u hrotů střel.

Ačkoliv je u antropologických dat nejčastěji používána metoda maximální parsimonie, používají se někdy i fenetické metody založené na vzdálenosti mezi jednotlivými analyzovanými objekty (tj. na podobnosti). Pak se většinou na základě buď binární matrix, nebo matrix, která umožňuje více variant, počítá vzdálenost mezi jednotlivými páry v rámci souboru. Je to potřebné například při použití metody Neighbor-Net (Coward et al.), program je ale schopný vzdálenosti spočítat sám. Cochrane a Lipo (2010) používali k výpočtu podobnosti jednotlivých taxonů Hammingovu vzdálenost, což je vlastně počet rozdílných znaků mezi 2 taxony, protože používali jinou fenetickou metodu- síť fenetických vzdáleností. Jejich práce ukazuje, že hlavním zdrojem inovací u oceánských kmenů byl horizontální transfer. Problémem jejich práce je ale právě použití fenetických metod. Pokud mají

jednotlivé populace podobné prvky díky kontaktu, v clusterové analýze převáží právě tyto podobnosti, bez ohledu na skutečnou fylogenezi.

Kromě kódování samotných analyzovaných dat je ještě někdy třeba vytvořit kódovaná data pro veličiny, s nimiž je předpokládána korelace. I to se obvykle dělá pomocí binárního kódování (např. Jordan & Shennan, 2003).

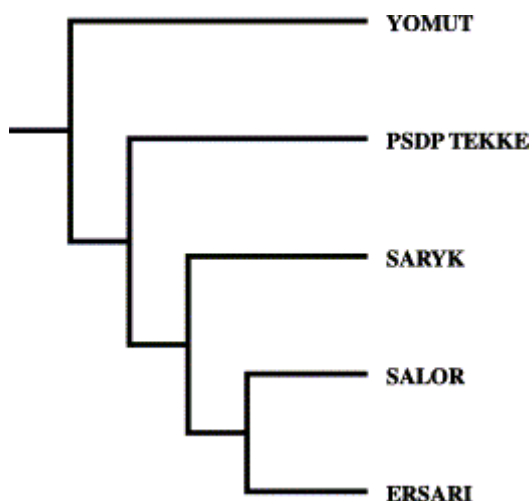
4.3 Analýzy a výsledky

Úspěch samotné analýzy pak závisí na vhodnosti zvolené metody. Studie v této oblasti nejčastěji pracují s metodou maximální parsimonie (O'Brien et al., 2001 Tehrani & Collard, 2002, Cochrane & Lipo, 2010 atd.). Dalšími používanými metodami jsou Neighbor-Net (Cochrane & Lipo, 2010) a Bayesiánská analýza (Matthews et al., 2011, Fortunato et al., 2006). Některé studie ještě používají další, nefylogenetické analýzy, když zkoumají kromě fylogeneze i něco navíc. Nejčastěji se práce zabývají možnou závislostí vývoje určitého prvku na jiných proměnných. Borgerhoff-Mulder et al. (2001) používají Concentrated Changes Test, který testuje možnou propojenost změn ve dvou binárně kódovaných proměnných. V novějších studiích se pro stejný test používá DISCRETE (Pagel, n.d.), což je vlastně test maximum-likelihood pro zjištění asociace binárních diskrétních dat.

Na konkrétních studiích, zaměřujících se na tkané textilie, je dobře vidět, jak tyto analýzy fungují. Tehrani a Collard (2002) ve své práci analyzovali vzor na turkmenských textiliích. Cílem jejich práce bylo jednak určit, jestli převažující vliv má v tomto případě fylogeneze nebo etnogeneze, a v případě, že převažuje fylogeneze, určit na základě analýzy vztahy mezi 5 skupinami Turkmenů (konkrétně Ersari, Salor, Saryk, Tekke a Yomut). Pracovali se 60 předměty, u kterých určili typické prvky a ty potom kodovali jako binární matrix. Výsledný soubor poté analyzovali dvěma způsoby. Použili PTP (permutation tail probability) test, který původně sloužil k určení toho, zda daná data obsahují fylogenetický signál. Tato metoda vytváří kladogramy z dat nepermutovaných a permutovaných. Signál se dá považovat za průkazný, pokud je délka kladogramu z nepermutovaných dat kratší než délka 95% stromů z dat permutovaných. Test v tomto případě fylogenetický signál prokázal. Druhou metodou byla maximální parsimonie a s ní spojené hodnocení stromů indexem konzistence a bootstrappingem. Zde vznikl jeden výsledný strom (obr.4). Ten je zároveň podpořen staršími

výsledky, jména klanů se dají rozdělit na 2 skupiny podle původu- Ersari, Saryk a Salor jsou oghuzská, zatímco Tekke i Yomut vykazují perský vliv. Dále výsledný strom podporuje geografické rozložení skupin. V rozporu s těmito důkazy jsou etnohistorické důkazy. Ty ovšem mohou být zkreslené, protože knihy z této doby ne vždy zaznamenávaly události věrohodně. Kromě kladogramu autoři také zjišťovali, která ze skupin nejvíce přispěla k etnogenezi skupiny Tekke. Po vyjmutí různých skupin z analýz autoři srovnali podporu bootstrapu a celkové kladogramy. Zjistili tak, že k etnogenezi v tomto případě nejvíc přispěl vzor Salor. Jako jedna ze starších prací tato studie tedy obhájí tvrzení, že fylogeneze hraje významnou roli při předávání kulturních znaků. Tuto hypotézu potvrdili Tehrani a Collard (2009) i později při analýze íránských tkaných textilií. Tyto studie jasně dokazují, že u některých kulturních znaků hraje fylogeneze minimálně stejně významnou roli jako horizontální transfer. V druhé práci ale výsledky podporoval index retence pouhých 0,59, což není výsledek, který by etnogenezi vylučoval tak moc, jak autoři naznačují a určitě je ještě potřeba se na problémy s velikostí jednotlivých indexů zaměřit.

Stejný přístup použili i O'Brien et al. (2001) v analýze vztahů mezi hroty střel. Kromě kladistického přístupu otestovali i přístup fenetický a ukázalo se, že i ten dobře funguje, což ale může ukazovat právě na korelace v datech (viz výše). Na tuto práci navázali Buchanan a Collard (2007) a analyzovali hroty z celých Spojených Států Amerických. Zároveň výsledný kladogram porovnávali se 4 ideálními kladogramy pro různé scénáře šíření populací. Problémem ale byly příliš malé rozdíly mezi vhodností jednotlivých scénářů.



Obrázek 4 Tehrani & Collard (2002)- výsledný kladogram turkmenských tkaných textilií

Problematictější je analýza hudebních nástrojů (Tëmkin & Eldredge, 2007), ve které autoři kritizují předchozí studie a zkouší metody použít na data, kde se fylogeneze nedá předpokládat jako jediný způsob vývoje. Analýza se potom neshoduje se známou fylogenezí. Problémem této analýzy může být i existence korelace mezi znaky, vzhledem k tomu, že velké množství jich bude spíše funkčních a budou spolu souviset. Navíc žádná z analýz nepředpokládala, že by se fylogenetické metody daly aplikovat na všechny lidské artefakty, takže práce v tomto ohledu nepřináší nic nového.

Použití Bayesiánské metody vyžaduje určité parametry. Toho využili k otestování hypotéz Matthews et al.(2011), kteří takto zjišťovali, jestli se vzory na íránských textilích chovají jako hierarchicky uspořádaný systém s určitou pevnější tradicí v jádře nebo jako systém skládající se z mnoha soudržných jednotek. Práce potvrdila druhou hypotézu. Jako jednu z metod používají Bayesiánskou analýzu ve své práci i Coward et al. (2008). Výsledky z ní se od metody maximální parsimonie liší v několika bodech, důležitější ale je, že tato práce ukazuje, že obě metody jsou schopné zaznamenat konflikt v datech, takže ani jeden z výsledných stromů nemá dostatečnou podporu, aby se dal brát jako definitivní výsledek. Autoři kromě těchto dvou metod použili ještě Neighbor-Net a tato metoda potvrdila spíše síťovité uspořádání a tedy to, že zkoumaná data se nedají určitě vysvětlit čistou fylogenezí. Neighbor-Net je ale metoda fenetická, viz výše, a nedá se z ní tedy vyvozovat fylogeneze, může jen ukázat, jak jsou si jednotlivé analyzované prvky podobné. Na tento fakt autoři často zapomínají a výsledné síť nebo stromy vytvořené clusterovacími metodami jako fylogenetické scénáře prezentují. Fortunato et al. (2006) pak tuto metodu využily i při rekonstrukci historie věna a naopak peněz, které muž musel zaplatit za ženu. Primárně je metodou vytvořen strom založený na jazykových skupinách a pak teprve jsou srovnány jednotlivé strategie. Z výsledného kladogramu je pak možné odhadnout ancestrální stav této strategie. Stejně přistupuje i k dalším analyzovaným kulturním znakům- monogamii a polygynii (Fortunato, 2011), i tomu, zda partneři zůstávají po svatbě žít v místě původu muže nebo ženy (Fortunato & Jordan, 2010). V obou studiích je problém řešen stejně, za použití jazykových stromů, v druhé jmenované jsou to kromě indoevropských jazyků ještě jazyky austronéské. Jordan et al. (2009) navíc už ve starší práci určili, že ancestrálním typem tohoto znaku bylo bydlení matrilocální, na základě doplnění jednotlivých strategií do fylogenetického stromu austronéských jazyků.

Dalším typem studií jsou práce, které zjišťují korelaci mezi některými kulturními znaky. Pro podobné práce je vždy nutné mít nějaký základní kladogram, na základě kterého budeme moci různé korelace testovat. Nejlepší by byla data genetická, ale ne u všech testovaných oblastí jsou k dispozici. Autoři tedy často jako náhradní řešení volí kladogramy sestavené pomocí dat lingvistických, někdy pomocí fylogenetických metod. V případě, že tyto nejsou k dispozici, stačí i výsledky srovnávací analýzy. Znaky spojené s manželstvím a celkovou organizací společnosti analyzovali Borgerhoff Mulder et al. (2001) a testovali různé možné korelace. Jako základ v této práci sloužil právě kladogram populací založený na lingvistických datech. Hypotézu, že společná fylogeneze tyto asociace neovlivňuje, se ale potvrdit nepodařilo, naopak ukázali, že i v regionálních studiích má fylogeneze na svědomí více než 60% asociací. Holden a Mace (2003) pracovaly s podobnými daty, jako základ použili strom jazyků Bantu (Holden, 2002). Jejich práce potvrzuje asociaci zániku matrilinearity a rozšíření chovu dobytka u bantuských kmenů. Naopak analýzou košíkářských výrobků kalifornských Indiánů se Jordan a Shennan (2003) snažili znovu, trochu jiným způsobem, dokázat roli fylogeneze v kulturní evoluci. Opět dochází k určování znaků na jednotlivých výrobcích, z nichž je posléze vytvořena binární matrix. Tato data pak autoři srovnávají s dalšími binárními matrix, vytvořenými z dat, které se týkají geografické vzdálenosti, ekologických podmínek a jazykových skupin. Pomocí metody, která se nazývá Mantel Matrix Test pak dostávají u každého páru korelační hodnotu každého páru. Tyto hodnoty ukazují, že nejsilněji koreluje s košíkářskými výrobky vzdálenost jednotlivých populací, ale zároveň také to, že pouze vzdálenost k vysvětlení variability nestačí. Na druhou stranu práce ukazuje i to, že na čím nižší úroveň v rámci populací jdeme, tím je pravděpodobnější, že rozmanitost košíkářských výrobků bude korelovat s jazykovou rozdílností dílčích populací. I tato analýza ale ukazuje nezbytnost mít na paměti možné korelace, protože pracuje i s různými technikami pletení košů, které ale mohou tvořit constraints pro další znaky a tím pádem snižovat počet znaků nezávislých.

I v rámci těchto analýz se objevilo několik analýz, které vytvářely umělá data, aby otestovaly spolehlivost jednotlivých metod. Nunn et al. (2010) vytvářejí vlastní umělá data, ve kterých pak dále simulují i extinkci, horizontální transfer a kolonizaci nového prostoru, takže dostávají umělý model evoluce určitých znaků. Snaží se poté zjistit, které údaje svědčí o fylogenezi. Vysoké hodnoty indexu retence a konzistence korelují s nízkou mírou horizontálního přenosu, ale nízké hodnoty naopak ne vždy prokazují etnogenezi, může se

jednat o jiné vlivy, což je nutné zohlednit v dalších analýzách, které často v podobných případech právě etnogenezi předpokládají. Autoři také používají dva kulturní znaky přenášené společně, v případě vývoje znaků na sobě nezávislých může být určování horizontálních transferů podstatně snazší.

5 Fylogenetická analýza lingvistických dat

5.1 Evoluce jazyků

Lingvistická data patří vlastně také mezi antropologická data, ale protože tvoří poměrně samostatnou a nejvíce analyzovanou skupinu, je jim v této práci věnován prostor v samostatné kapitole. Jazyky jsou stejně jako ostatní kulturní znaky přenášeny z generace na generaci a při odštěpování jednotlivých populací může docházet ke kladogenezi. Zároveň ale samozřejmě probíhá i anageneze a syngeneze (vznik kreolštin). U jazyků se tedy dá přepokládat silná fylogeneze, ale je zároveň nutné si uvědomit, že i v rámci jazyků dochází k horizontálnímu transferu, konkrétně k jazykovým výpůjčkám. Například u angličtiny se někdy uvádí, že minimálně 40% slovní zásoby jsou výpůjčky z francouzštiny z doby vlády Viléma Dobývatele.

Že jsou mezi sebou jazyky příbuzné, si lingvisté začali uvědomovat poměrně brzo a používání stromů, které tuto příbuznost měly ukazovat, bylo ještě čtenější, než v tehdejší biologii (Atkinson & Gray, 2005). Zároveň ale pro konstrukci stromů používali jiné znaky a nejednalo se tedy o stromy fylogenetické. Na základě srovnávacích analýz používali především synplesiomorfie (tedy prvky společné taxonům odvozeným i primitivním), ať už na úrovni fonetické nebo lexikální. Ve 20. století došlo i k vývoji počítačové metody, která měla sloužit k usnadnění analýzy. Lexikostatistika sloužila k určení vzdálenosti mezi jazyky na základě sdílených kognátů a konstrukci stromů pomocí clusterovacích technik. K tomu se používala UPGMA (viz. 2.2). Později byla ještě vyvinuta glottochronologie, která pracovala s představou, že existuje konstantní rychlost, s jakou se lexikon obměňuje (Atkinson & Gray, 2005). Obě tyto metody jsou ale dnes spíše kritizovány, protože se jedná o metody nefylogenetické a navíc se ukazuje, že výsledky byly dost nepřesné (Barbançon et al., 2006).

Na začátku 21. století ale došlo k použití fylogenetických, primárně biologických metod na lingvistická data a od té doby bylo publikováno velké množství studií, které se zabývají nejrozličnějšími jazykovými rodinami, používají různé typy dat a různé metody a na základě výsledného kladogramu se snaží podpořit jiné hypotézy, spojené s historickým vývojem lidské populace. Stejně tak se ale používají i metody nefylogenetické (SplitsTree a Neighbor-Joining) a v poslední době se objevují i studie používající síť.

5.2 Výběr a kódování dat

Při výběru dat je nutné si stále uvědomovat, že mezi jazyky probíhá horizontální transfer a výběr tomu přizpůsobit. V současnosti se používají především data lexikální, ale v některých pracích autoři použili i data jiná- morfologická (Barbaçon et al., 2006), fonologická (Atkinson, 2011) nebo oboje (Rexová et al., 2006). Zároveň i v rámci jednotlivých kategorií je nutné dále vybírat. Při použití lexikálních dat se používají kognáty, což jsou slova stejného původu, ačkoliv tak na první pohled nemusí působit. Kognáty jsou například anglické slovo *mother*, německé *Mutter* a italské *madre*, která všechna pocházejí z předpokládaného indoevropského slova **mater*. Kromě toho je dobré vybrat si slova, u kterých je nižší pravděpodobnost, že se do jazyka budou dostávat jako výpůjčka z jazyka jiného. To se týká obvykle slov, která jsou často používaná a označují základní věci, jako jsou členové rodiny, číslovky, označení pro základní skutečnosti jako *muž*, *žena*, *jazyk*. U takovýchto výrazů je pak podstatně nižší pravděpodobnost, že výsledek bude ovlivněn příliš rozsáhlým horizontálním transferem.

Obvykle se také používají už dříve sebraná slova, kde už jsou kognáty určené- Holden (2002) použila soubor kognátů jazyků Bantu (celkem 92 významů v 87 jazycích) ze starší, lexikostatistické práce Bastin et al. (1999), stejně jako později Rexová et al. (2006). Tento seznam navíc vychází z listu 100 základních slov, který vytvořil Swadesh (1955). V případě indoevropských jazyků se používá lexikostatistický dataset, který vytvořili Dyen et al. (1992) a který obsahuje celkem 200 různých významů v 84 jazycích. V pracích o austronéských jazycích se pak někdy používají ještě větší soubory dat- 400 jazyků a 210 významů v Greenhill & Gray (2009), což jsou zároveň autoři *Austronesian Basic Vocabulary Database*, se kterou pracují i ve svých dalších pracích. Někteří autoři pracovali i s historickými jazyky,

Forster & Toth (2003) vytvořili soubor 35 znaků, do kterého zahrnuli i staré jazyky jako jsou gaelština a keltské jazyky. Některé doklady pak získávali z latinsko- galských nápisu.

Kromě lexikálních dat některé práce používají data typologická. Dunn et al. (2005) pracovali se 123 typologickými prvky v jazycích austronéských a papuánských. Jazyková typologie pracuje s několika ideálními jazykovými typy, které mají určité rysy. Tyto rysy jsou někdy spojené a týkají se takových věcí, jako je slovosled, flexe (ohýbání tvarů slov- vyjádření různých mluvnických kategorií) apod. Například u jazyku, který je flektivní (jako čeština) je vysoká pravděpodobnost, že slovosled bude poměrně volný, protože vztahy mezi větnými členy bude možné vyjádřit právě flexí. Naopak angličtina, jako jazyk analytický a na flexi chudý, bude mít slovosled pevně daný (podmět-sloveso-předmět). Problémem u takovýchto dat pak může být opět korelace některých znaků, kterou Dunn et al. (2005) ve své práci neřeší. Stejně tak se tomuto problému nevěnuje ani Bakker et al. (2011), který stejnou metodou zkoumá vztahy mezi kreolštinami a jazyky nekreolskými. Jeho práce ale používá pouze metody clusterovací a po fylogenezi nepátrá, takže korelace některých znaků by neměla výsledek výrazněji ovlivnit.

Data morfologická a fonologická (tedy gramatická) se obecně považují za pomaleji se měnící prvky jazyka. Pokud jim ale dodáme váhu, tak už použitá metoda nemůže být striktně kladistická (Barbançon et al., 2006). Proto se obvykle používají data nevážená. Navíc u základních významů se obvykle daří ze souboru vyloučit výpůjčky a tím pádem omezit homoplázie, které by ovlivnily výsledný strom.

Data se potom kódují jako u předchozích analýz, nejčastěji jako binární matrix. Pro jednotlivé kognáty se u jednotlivých jazyků kóduje jejich přítomnost/nepřítomnost. Některé analýzy (např. Holden, 2002) povolují polymorfismus, vzhledem k tomu, že pro jeden význam mohou mít jazyky více slov. V tomto případě ale může být problém s přesným sémantickým vymezením slova, protože absolutní synonyma, tedy slova se stejným významem, se v jazycích prakticky nevyskytují. Kromě binární matrix bývá ještě někdy používána matrix, která umožňuje větší počet variant a pak každá varianta koresponduje s jednou skupinou kognátů (Rexová et al., 2002).

Kromě metod založených na diskrétních datech se používají i metody založené na vzdálenostech a pak se obvykle opět počítá vzdálenost jako vzdálenost mezi dvěma taxony na základě rozdílů.

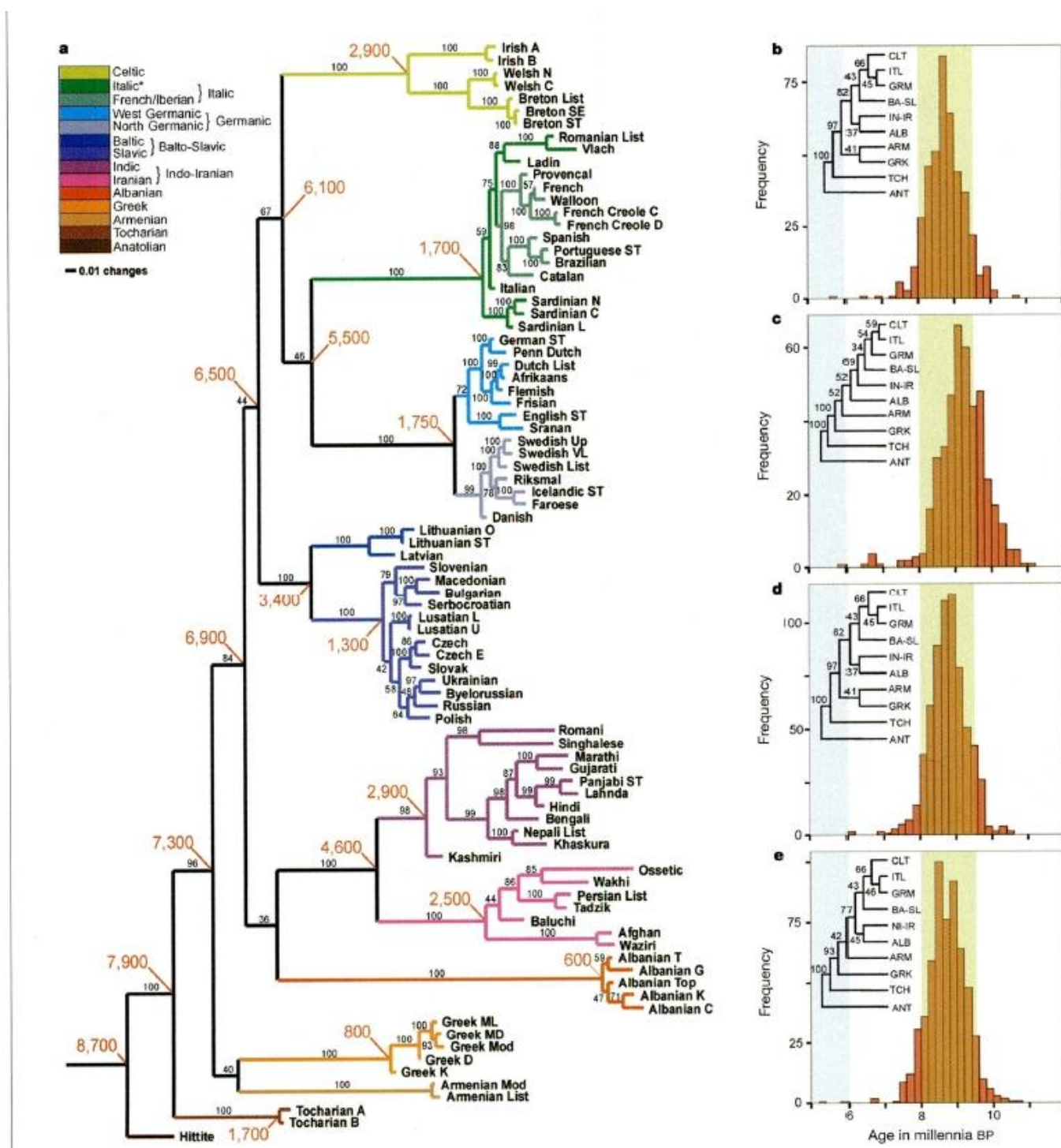
5.3 Metody a analýzy

5.3.1 Indoevropská jazyková rodina

Indoevropská jazyková rodina je nejlépe prozkoumanou jazykovou rodinou na světě, existují pro ni mnoho srovnávacích analýz, a proto se v začátku aplikace fylogenetických metod přímo nabízela, protože data jsou dostupná, kognáty určené a pro porovnání k dispozici výsledky komparativní analýzy. Je proto také v rámci fylogenetických analýz rodinou nejstudovanější. Byly provedeny studie, které určovaly její fylogenezi (Gray & Atkinson, 2003; Rexová et al., 2003), studie snažící se určit stáří rodiny (Atkinson & Gray, 2006) i studie, které pracovali s frekvencí různých slovních druhů (Pagel et al., 2007)

Ačkoliv je tato jazyková rodina jednou z nejprozkoumanějších, umístění některých skupin stále není zcela jasné. Dvě práce konstruující fylogenetické stromy vyšly téměř současně a obě srovnávaly dvě hypotézy o původu Indoevropanů. Gray & Atkinson (2003) používají Bayesiánskou analýzu a na základě stromu (obr. 5) se vyslovují pro teorii, podle které Indoevropané přišli z Anatólie. Oproti starší analýze austronéských jazyků má tento strom index konzistence vyšší, ale pořád ne zcela spolehlivý. Druhou z nich byla práce Rexové et al. (2003), ve které se autoři nakonec nevyslovili pro žádnou z hypotéz, protože u některých důležitých skupin stále panovala nejistota, kam by na stromě měly patřit (albánština a germánské jazyky). Jak je vidět na obr.6, výsledky obou analýz se od sebe mírně liší, což může být způsobeno odlišnou metodikou obou prací. To jen zdůrazňuje, jak důležité je jednotlivé metody dobře zanalyzovat a otestovat. Nejnovější práce z této oblasti se vrací k metodě UPGMA, i když vzdálenosti mezi taxony počítá jinak (Serva & Petroni, 2008). Překvapivě se ale její výsledek shoduje se stromem b z obr.6. Vzhledem k tomu, že metoda použitá v této práci nepotřebuje určení kognátů, určitě stojí za bližší prozkoumání, protože u některých jazykových skupin by takové určení bylo velmi náročné a tato metody by analýzu

mohla velmi usnadnit.

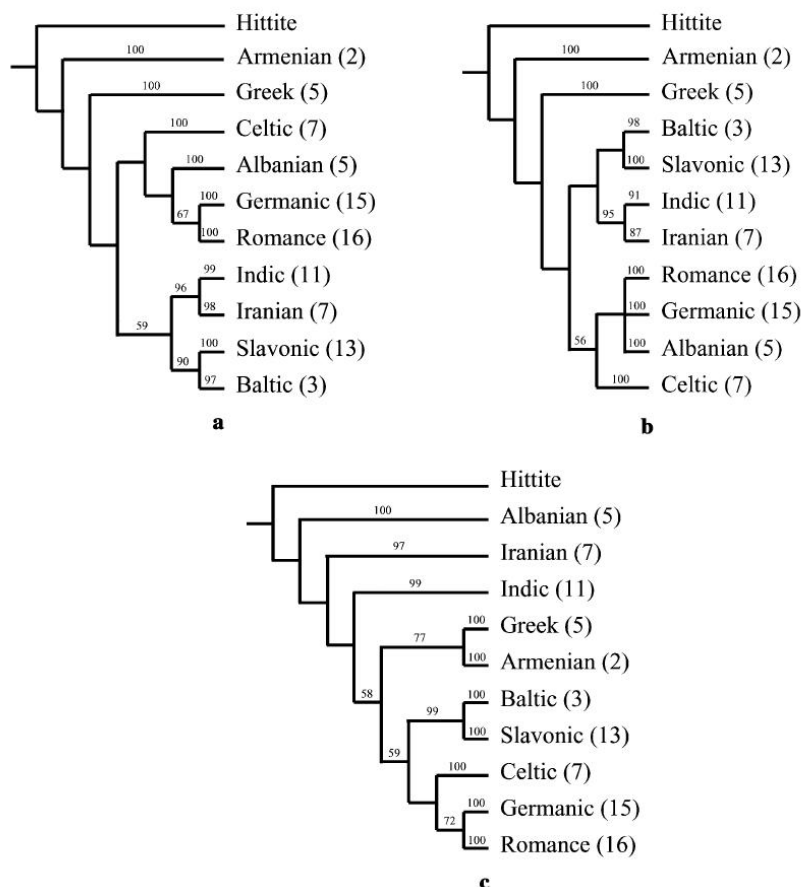


Obrázek 5 Gray & Jordan (2003)- fylogenetický strom indoevropských jazyků

Kromě analýzy fylogeneze se u indoevropských jazyků zkoušely ještě dvě zajímavé metody. Pagel et al. (2007) ukazují, že jazyková evoluce má určitá pravidla a že různé slovní druhy se vyvíjejí různou rychlostí. Vzhledem k tomu, že frekvence použití slov koreluje

s rychlostí jejich evoluce, dá se předpokládat, že některá slova jsou podstatně starší, než se čekalo, ačkoliv jazyková evoluce probíhá relativně rychle.

K. Rexová et al. / Cladistics 19 (2003) 120–127



Obrázek 6 Rexová et al. (2003)- 3 stromy indoevropských jazyků, v závislosti na způsobu kódování dat

Poslední výzvou je potom rekonstrukce stáří celé jazykové rodiny. Glottochronologie sice stáří byla schopná určit, ale vycházela z chybného předpokladu, že evoluce probíhá vždy stejně rychle. Dnes pouze některé metody tvoří stromy, kde délka větví odpovídá počtu změn v evoluci a tedy často i délce evoluce. Jednou z nich je i Bayesiánská metoda, se kterou pracují Gray, Atkinson, Greenhill atd. ve většině svých prací. Atkinson a Gray (2006) tak stanovili stáří této rodiny na 8000-9000 let. Problémem této analýzy ovšem je, že vyžaduje určení stáří některých uzlů, aby se ke konkrétnímu datu dostala, a to nikdy s jistotou nevíme.

5.3.2 Austronéska jazyková rodina

Austronéska jazyková rodina je jednou z vůbec největších. Austronéskými jazyky se hovoří v Oceánii, na Novém Zélandu, na ostrovech jihovýchodní Asie a na Madagaskaru. Je

to vůbec první jazyková rodina, která byla analyzována pomocí kladistiky (Gray & Jordan, 2000). První práce na základě fylogenetického stromu zkoumala hypotézu o kolonizaci Pacifiku. Scénář „express train to Polynesia“ předpokládá velice rychlé rozšíření po Pacifiku (obr. 7) a pokud platí, mělo by pořadí odvětvování jazyků odpovídat datům archeologickým. Autoři použili poměrně velký dataset a je možné, že neidentifikovali správně všechny výpůjčky, což by vysvětlovalo relativně nízký index konzistence 0,25. Autoři sami tvrdí, že by to mohlo být způsobené jiným charakterem dat, než mají data biologická, ale tento předpoklad novější studie nepotvrzují. I tak ale data tvoří strom (obr.8), který je konzistentní s výše zmiňovaným scénářem (ačkoliv ne všechna větvení budou pravděpodobně správná, je na něm vidět postupné odštěpování větví, konzistentní s migrací bez dalšího kontaktu a bez vracení se zpět. Umožňuje tak odpovědět na námitku obhájců modelu, který předpokládá spíše rozsáhlejší kontakt mezi jednotlivými populacemi a nemožnost rekonstruovat kulturu pomocí fylogenetických metod.

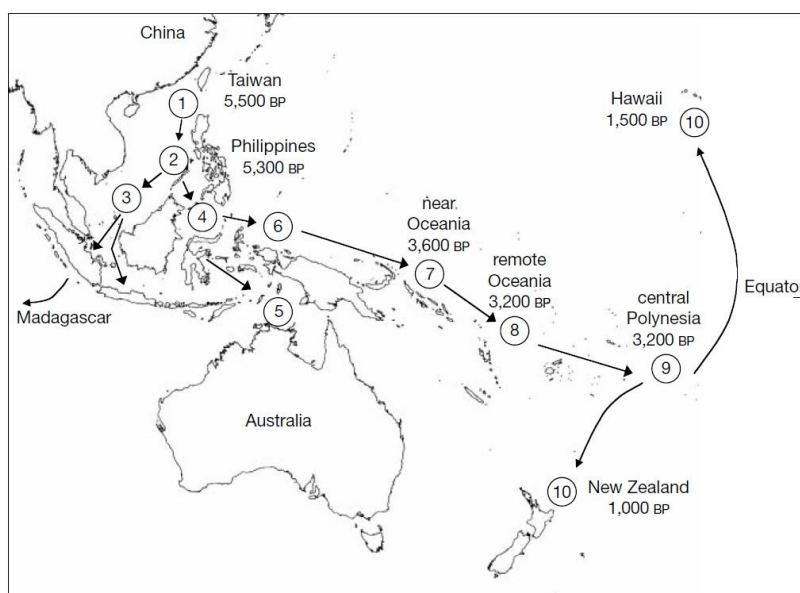


Figure 1 The express train to Polynesia model of the Austronesian colonization of the Pacific (adapted from refs 5 and 22). Approximate archaeological dates of settlement are indicated^{13,22}. Each 'station' is a separate character state: 1, Taiwan; 2, Philippines,

Chamorro, Palau; 3, Borneo, Indonesia, Malay; 4, Sulawesi; 5, central Malayo-Polynesian; 6, south Halmahera/west New Guinea; 7, near Oceania; 8, remote Oceania; 9, central Polynesia; 10, east Polynesia.

Obrázek 7 Gray & Jordan (2000)- model expanze Austronesanů z Taiwanu

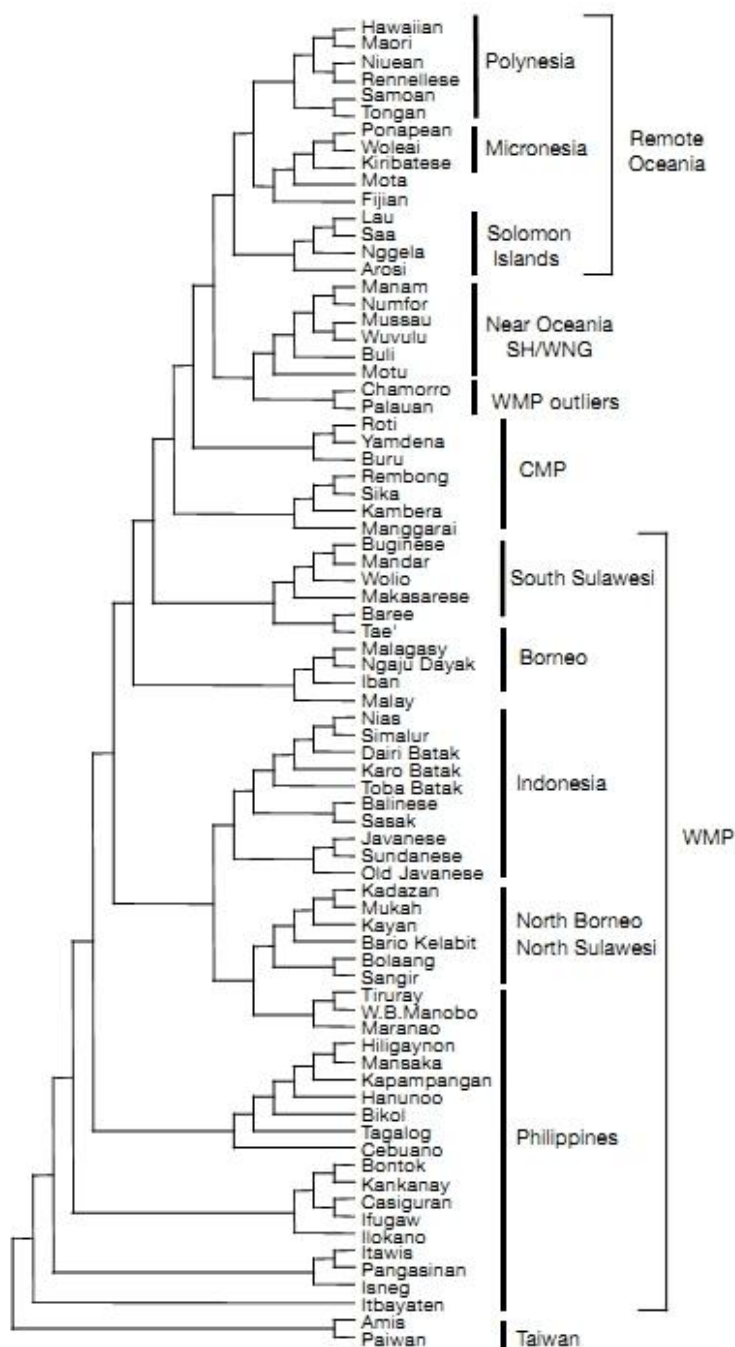


Figure 3 Phylogenetic tree of 77 Austronesian languages. WMP, Western Malayo-Polynesian; CMP, Central Malayo-Polynesian; SH/WNG, South Halmahera/West New Guinea. The topology of the tree shows considerable agreement with traditional linguistic groupings^{14,15,28}; these groupings reflect historical relationships, not just geographical proximity. For instance, Malagasy (spoken on Madagascar) is grouped with Ngaju Dayak from western Borneo. Tae' (from Central Sulawesi) groups within the south Sulawesi languages, whereas the north Sulawesi languages (for example, Bolaang Mongondow (Bolaang)) are more closely related to languages of north Borneo than are other Sulawesi languages.

Obrázek 8 Gray & Jordan (2000) - fylogenetický strom austronéských jazyků

Kromě ověřování některých hypotéz ale je ale také snaha využít nové metody na analýzu dat, která byla dosud obtížná. Dunn et al. (2005) analyzují pomocí Bayesiánské metody kromě jazyků austronéských i papuánské. Používají k tomu typologická data, protože v současných papuánských jazycích není možné najít dostatek kognátů. Toto jen potvrzuje, že evoluce jazyků probíhá rychleji než evoluce biologická, a že se papuánské jazyky musely oddělit relativně brzo. Použití typologických dat bylo otestováno na menším souboru se známou fylogenezí a teprve potom aplikováno na zkoumaný soubor. Problémem u typologických dat ale může být korelace (viz výše)

5.3.3 Jazyky Bantu

Bantuské jazyky jsou jazyky, kterými se mluví v téměř celé jižní Africe a v některých oblastech Afriky východní a střední. Patří do Nigerokonžské jazykové rodiny. Jsou také hned po austronéských jazycích první skupinou, na kterou byla kladistická analýza použita. Data jsou totiž dobře dostupná, alespoň co se základní slovní zásoby týče, a navíc existují hypotézy o migraci jednotlivých kmenů, které by tato analýza mohla podporovat nebo popírat.

Holden (2002) provedla kladistickou analýzu bantuských jazyků, ve které použila 92 lexikálních významů. Použila metodu maximální parsimonie. Její analýza ukazuje korelaci mezi lexikálními a archeologickými daty a podporuje tak hypotézu o rozšíření kmenů v průběhu rozšíření zemědělství. Rexová et al. (2006) pracují s větším množstvím znaků než Holden, protože kromě lexikálních dat používají i data gramatická. Byl použit set 92 lexikálních významů a 52 fonologických a morfologických znaků (Bastin et al., 1979). Zároveň tato analýza porovnává metodu maximální parsimonie a Bayesiánskou analýzu a ukazuje, že výsledky těchto dvou metod se od sebe příliš neliší. Naopak se tato analýza liší od první kladistické větším počtem jazyků a větším počtem znaků. Je tedy vidět, že počet použitých znaků analýzu ovlivňuje a při použití příliš malého datového souboru může výsledky zkreslovat. Analýza, kterou provedla Holden, má ale vyšší indexy konzistence i retence, což může být způsobeno právě výběrem datasetu. Ve větším souboru se může vyskytovat více výpůjček. Také by to mohlo naznačovat, že i morfologické a fonologické struktury mohou být ovlivněny kontaktem.

Tyto analýzy byly také použity ke korelaci s některými antropologickými daty v práci Holden & Mace (2003), viz výše.

6 Závěr a diskuze

Fylogenetické metody se v poslední době aplikují nejen na data biologická. Donedávna mnoho vědců tvrdilo, že kulturní evoluce probíhá jinak než evoluce biologická a že klíčový je v ní kontakt mezi jednotlivými kulturami, zatímco fylogeneze tak velkou roli hrát nemůže (Mesoudi et al., 2004). Dnes se ukazuje, že fylogeneze, ačkoliv není jediným procesem, který způsobuje variabilitu mezi kulturami, je v kulturní evoluci důležitá a dá se tak vysvětlit velké množství variability. Samozřejmě někde dochází k extenzivnímu horizontálnímu transferu a pak se vztahy mezi danými taxony nedají vyjádřit pouze fylogenetickým stromem. U velkého množství fylogenetických analýz kulturních dat ale výsledky souhlasí s výsledky analýz prováděných jinými metodami a potvrzují tak, že aplikace nových metod má smysl. Oblasti, ve kterých se metody používají, jsou především analýzy příbuznosti rukopisů, analýzy antropologických dat a analýzy lingvistické. Ve všech oblastech je ale stále potřeba jistá opatrnost, protože zdaleka ne všechno data se dají použít a zdaleka ne všechny metody se dají aplikovat se stejným úspěchem.

Analýzy rukopisů jsou nejstarší oblastí aplikace těchto metod. Ukazuje se zde, že podobnost rukopisů málokdy vzniká horizontálním transferem a proto zde dobré výsledky podávají i metody nefylogenetické. Ačkoliv metody ještě nejsou zcela ideální a někdy může být nutná podrobnější analýza některých částí ručně, stále je to způsob jak urychlit a zjednodušit práci a při dodržení určitého postupu také vyřadit z analýzy subjektivní lidský faktor. Celkově je aplikovatelnost těchto analýz omezena hlavně dostupností dat, sehnat všechny rukopisy určitého textu, které ještě nejsou digitalizované, je hlavní komplikací této práce. I proto autoři analyzují hlavně kanonická díla, jako jsou *The Canterbury Tales*, *Monarcha* nebo kratší texty, u kterých je digitalizace méně náročná.

Oblast antropologických dat by do sebe vlastně mohla zahrnout obě ostatní oblasti, ale v této práci sem řadím především hmotné artefakty a sociální uspořádání v některých kulturách. I zde se výsledky porovnávají se staršími výzkumy. Stanovovat příbuznost pouze na základě jednoho typu hmotného artefaktu je metoda, kterou používá právě až fylogenetická

analýza, ale ta se dá srovnat s výzkumy založenými na jiných datech. Navíc je dost často možné použít pro srovnání právě data lingvistická, ať už z komparativní nebo fylogenetické analýzy. Takové analýzy se tedy hodí pro zkoumání korelace jednotlivých proměnných. Dá se takto i rekonstruovat ancestrální stav jednotlivých znaků, což bylo prozatím využito především u analýzy různých společenských systémů (věno x cena za nevěstu, monogamie x polygynie atd.). Stejně tak se dá srovnat korelace vývoje určitého kulturního znaku s různými znaky dalšími (geografické rozložení, populace definované pomocí jazyků, ekologie dané oblasti) a pak je možné zjišťovat, zda v případě konkrétního znaku převládá etnogeneze nebo fylogeneze. Cílem dalších prací by určitě mělo být analyzovat nová data (v současnosti se opakují analýzy dat podobných- 4 práce zabývající se hroty střel, 4 práce zabývající se tkanými textiliemi atd.). Pravdou ale je, že taková data nebude snadné vybrat. Zároveň by se budoucí práce měly soustředit i na rozdíly mezi vertikálním a horizontálním transferem (tedy etnogenezí a fylogenezí) a snažit se zjistit, kterými faktory je ovlivněna převaha jedné nebo druhé možnosti.

Fylogenetická analýza jazyků je největší z předchozích oblastí. Je to dáno množstvím a rozmanitostí prací, které se jazykům věnovaly. První byly studie, které na základě rekonstrukce fylogeneze jazykových rodin potvrzovaly nebo vyvracely hypotézy týkající se migrace a původu konkrétních populací. Na začátku se také pracovalo především s daty lexikálními, v současnosti už jsou ale běžně využívána i data jiná. Jednotlivé práce už se také nesoustředí jen na konkrétní analýzy, ale objevuje se velké množství studií hodnotících jednotlivé metody. Práce rekonstruující fylogenezi jazykových rodin se zároveň snaží určit stáří jednotlivých kladů. Tato snaha je ale stále nepříliš přesvědčivá, protože rozpětí dat, které současné metody poskytují, je příliš velké. Navíc je pro takovou analýzu potřeba určit alespoň několik constraints v místech, kde přibližné stáří známe, což může být v některých případech zavádějící. Stejně tak snaha určit na základě stáří jazyků stáří jednotlivých populací je v současnosti stále nepřesná, protože pokusy o sloučení dat genetických a lingvistických nemusí fungovat, vzhledem k obrovským rozdílům v rychlosti vývoje. Jazyky se vyvíjejí podstatně rychleji. Dostupnost dat pro analýzu jazyků může být také někdy problematická, a to především u menších nebo méně populárních jazykových skupin. Tato práce ale zároveň dokazuje, že neprobádané jazykové skupiny stále existují a určitě by mělo význam aplikovat analýzy i na ně (to se týká například různých dialektů romštiny). Stejně tak jsou u některých analýz výsledky takové, že by určitě stálo za to je ověřit (semito-hamitské jazyky).

Celkově se tato práce pokusila shrnout výsledky ve třech velkých oblastech, ve kterých se používají fylogenetické analýzy a ukázat, že jejich použití má své opodstatnění, i když je nutné uvědomovat si i všechna omezení. Vychází z předpokladu, že jazyk jako fenotypový projev člověka nám může poskytnout další cenné informace o vývoji lidstva. Přípravuje tak prostor pro případné navázání vlastní analýzou.

7 Použitá literatura

- Atkinson, QD., Gray, RD. (2005) Curious parallels and curious connections- phylogenetic thinking in biology and historical linguistics. *Systematic Biology* 54: 513-526.
- Atkinson, QD., Gray, RD. (2006) How old is the Indo-European language family? Illumination or more moths to the flame?. In: Forster, P., Renfrew, C. *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: MacDonald Institute Press, 91–110.
- Atkinson, QD. (2011) Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332: 346-349.
- Bakker, P., Daval-Markussen, A., Parkvall M., Plag I. (2011) Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26: 5-42.
- Barbançon, F., Warnow, T., Evans, SN., Ringe, DA., Nakhleh L. (2006) An experimental study comparing linguistic phylogenetic reconstruction methods. Proceedings of a conference on Language and Genes, University of California, Santa Barbara, September 2006, 1-26.
- Barbrook AC., Howe, CJ., Blake N., Robinson, P. (1998) The phylogeny of *The Canterbury Tales*. *Nature* 394: 839.
- Baret, PV., Macé, C., Robinson, P. (2006) Testing methods on an artificially created textual tradition. In: Macé, C., Baret, PA., Bozzi A., Cignoni L. (eds) *The Evolution of Texts: Confronting Stemmatological and Genetical Methods*. *Linguistica Computazionale*, Vol. XXIV–XXV, Pisa: Accademia editoriale, 255–83.
- Bastin, Y., Coupez, A., Mann, M. (1999) Continuity and divergence in the Bantu languages: perspectives from a lexicostatistic study. *Annales, Sciences humaines*,: **162**.
- Borgerhoff Mulder, M., George-Cramer, M., Eshleman, J., Ortolan, A. (2001) A study of East African kinship and marriage using phylogenetically-based comparative methods. *American Anthropologist* 103: 1059–1082.
- Bryant, D., Moulton, V. (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic network. *Molecular Biology and Evolution* 21: 255–65.
- Buchanan, B., Collard, M. (2007) Investigating the peopling of North America through cladistic analyses of Early Paleoindian projectile points. *Journal of Anthropological Archaeology* 26: 366-393.
- Cochrane, EE., Lipo CP. (2010) Phylogenetic analyses of Lapita decoration do not support branching evolution or regional population structure during colonization of Remote Oceania. *Philosophical Transactions of the Royal Society B* 365: 3889–3902.
- Coward, F., Shennan, SJ., Colledge, S., Conolly, J., Collard M. (2008) The spread of Neolithic plant economies from the Near East to Northwest Europe: a phylogenetic analysis. *Journal of Archaeological Science* 35: 42–56.
- Darwin, C. (1859) *The Origin of Species by Means of Natural Selection*. Oxford: Oxford University Press.
- Dunn, M., Terrill, A., Reesink, G., Foley, RA., Levinson, SC. (2005) Structural phylogenetics and reconstruction of ancient language history. *Nature* 309: 2072-2075.
- Dyen, I., Kruskal, JB., Black, P. (1992) An Indoeuropean classification: a lexicostatistical experiment. *Transactions of American Philosophical Society* 82: 1– 132.

- Eagleton, C., Spencer, M. (2006) Copying and conflation in Geoffrey Chaucer's *Treatise on the astrolabe*: a stemmatic analysis using phylogenetic software. *Studies in History and Philosophy of Science* 37: 237-268.
- Forster, P., Toth A. (2003) Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European. *PNAS* 100: 9079-9084.
- Fortunato, L., Holden, C., Mace, R. (2006) From bridewealth to dowry? A Bayesian estimation of ancestral states of marriage transfers in Indo-European group. *Human Nature* 17: 355-376.
- Fortunato, L., Jordan, F. (2010) Your place or mine? A phylogenetic comparative analysis of marital residence in Indo-European and Austronesian societies. *Philosophical Transactions of the Royal Society B* 365: 3913-3922.
- Fortunato, L. (2011) Reconstructing the history of marriage strategies in Indo-European-speaking societies: monogamy and polygyny. *Human Biology* 83: 87-105.
- Fortunato, L. (2011) Reconstructing the history of residence strategies in Indo-European-speaking societies: Neo-, uxori-, and virilocality. *Human Biology* 83: 107-128.
- Gray, RD, Jordan, FM. (2000) Language trees support the express-train sequence of Austronesian expansion. *Nature* 405: 1052-1055.
- Gray, RD, Atkinson QD. (2003) Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426: 435-439.
- Greenhill, SJ., Gray, RD. (2009) Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods. In: Adelaar, A., Pawley, A. (eds.) *Austronesian Historical Linguistics and Culture History: a Festschrift for Robert Blust*. Canberra: Pacific Linguistics, 375-397.
- Harvey, PH., Pagel, MD. (1991) *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.
- Hennig, W. (1965) Phylogenetic Systematics. *Annual Review of Entomology* 10: 97-116.
- Holden, CJ. (2002) Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proceedings of the Royal Society of London B* 269: 793-799.
- Holden, CJ., Mace, R. (2003) Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proceedings of the Royal Society B* 270: 2425-2433.
- Howe, CJ., Barbrook, AC., Spencer, M., Robinson, P., Bordalejo, B., Mooney LR. (2001) Manuscript evolution. *Trends in Genetics* 17: 147-152.
- Huson, DH. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68-73.
- Jordan, FM., Gray, RD., Greenhill, SJ., Mace, R. (2009) Matrilocal Residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B*. published online <<http://rspb.royalsocietypublishing.org/content/early/2009/02/27/rspb.2009.0088.full.pdf+html>>
- Jordan, P., Shennan, S. (2003) Cultural transmission, language, and basketry traditions amongst the Californian Indians. *Journal of Anthropological Archaeology* 22: 42-74.
- Lantin, AC., Baret, PV., and Macé, C. (2004) Phylogenetic analysis of Gregory of Nazianzus' Homily 27. In: Purnelle, G., Fairon, C., Dister, A. (eds) *Le poids des mots: Proceedings of the 7th International Conference on the Statistical Analysis of Textual Data*. Louvain-la-Neuve: Presses Universitaires de Louvain, 700-707.

- Lee, AR. (1989) Numerical taxonomy revisited: John Griffith, cladistic analysis and St. Augustine's *Quaestiones in Heptateuchum*. *Studia Patristica* 20: 24–32.
- Matthews, LJ., Tehrani, JJ., Jordan, F., Collard, M., Nunn, CN. (2011) Testing for divergent transmission histories among cultural characters: a study using Bayesian phylogenetic methods and Iranian tribal textile data. *PLoS* 6: 1-10.
- Mesoudi, A., Whiten, A., Laland, KN. (2004) Perspective: Is human cultural evolution Darwinian? Evidence reviewed from the perspective of *The Origin of Species*. *Evolution* 58: 1-11.
- Nakleh, L., Ringe, D., Warnow, T. (2005) Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages. *Language* 81: 382-420.
- Nichols, J., Warnow, T. (2008) Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2: 760–820.
- Nunn, CL., Arnold, C., Matthews, L., Borgerhoff Mulder, M. (2010) Simulating trait evolution for cross-cultural comparison. *Philosophical Transactions of the Royal Society B* 365: 3807-3819.
- O'Brien, MJ., Darwent, J., Lyman, RL. (2001) Cladistics is useful for reconstructing archaeological phylogenies: Palaeoindian points from the southeastern United States. *Journal of Archaeological Science* 28: 1115–36.
- O'Brien, MJ., Lyman, RL., Mesoudi, A., VanPool, TL. (2010) Cultural traits as units of analysis- *Philosophical Transactions of the Royal Society B* 365: 3797–3806.
- Pagel, M. n.d. *User's Manual for Discrete*. School of Animal and Microbial Sciences. University of Reading.
- Pagel, M., Atkinson, QD., Meade, A. (2007) Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449: 717-721.
- Phillips-Rodriguez, WJ., Howe, CJ., Windram, HF. (2009) Chi-squares and the phenomenon of "change of exemplar" in the *Dyūtaparvan*. *Lecture Notes in Artificial Intelligence, Sanskrit Computational Linguistics*, Heidelberg: Springer, 380-390.
- Platnick, NI., Cameron, HD. (1977) Cladistic methods in textual, linguistic, and phylogenetic analysis. *Systematic Zoology* 26: 380–85.
- Rexová, K., Frynta, D., Zrzavý J. (2003) Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics* 19: 120-127.
- Rexová, K., Bastin, Y., Frynta, D. (2006) Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. *Naturwissenschaften* 93: 189-194.
- Robinson, PMW., O'Hara, RJ. (1996) Cladistic analysis of an Old Norse manuscript tradition. In: Hockey, S., Ide, N. (eds) *Research in Humanities Computing* 4, 115–37.
- Saitou N., Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
- Serva, M., Petroni, F. (2008) Indo-european tree by Levenshtein distance. *EPL* 81: 1-5.
- Skelton, C. (2008) Methods of using phylogenetic systematics to reconstruct the history of the linear B skript. *Archaeometry* 50: 158-176.
- Sokal, RR., Michener CD. (1958) A statistical method for evaluating systematic relationships. *The University of Kansas Scientific Bulletin* 38: 1409-1438.
- Spencer, M. Howe, CJ. 2001. Estimating distances between manuscripts based on copying errors. *Literary and Linguistic Computing* 16: 467-484.
- Spencer, M., Bordalejo, ., Wang, LS., Barbrook, AC., Mooney, LR., Robinson, P., Warnow, T., Howe, CJ. (2003) Analyzing the order of I'tems in manuscripts of *The Canterbury Tales*. *Computers and the Humanities* 37: 97-109.

- Spencer, M., Davidson, EA., Barbrook, AC., Howe, CJ. (2004) Phylogenetics of artificial manuscripts', *Journal of Theoretical Biology* 227, 503-511.
- Swadesh M. (1955) Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21:121–137.
- Swofford, D.L., (1996) PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods). Sunderland, Massachusetts: Sinauer Associates. Version 4.0.
- Tehrani, JJ., Collard, M. (2002) Investigating cultural evolution through biological phylogenetic analyses of Turkmen textiles. *Journal of Anthropological Archaeology* 21: 443-463.
- Tehrani, JJ., Collard, M. (2009) On the relationship between interindividual cultural transmission and population-level cultural diversity: a case study of weaving in Iranian tribal populations. *Evolution and Human Behaviour* 30: 286-300.
- Tehrani, JJ., Collard, M., Shennan, SJ. (2010) The cophylogeny of populations and cultures: reconstructing the evolution of Iranian tribal craft traditions using trees and jungles. *Philosophical Transactions of the Royal Society B* 365: 3865–3874.
- Tëmkin, I., Eldredge, N. (2007) Phylogenetics and material cultural evolution. *Current Anthropology* 48: 146-153.
- Windram, HF., Howe, CJ., Spencer, M. (2005) The identification of exemplar change in the Wife of Bath's Prologue using the maximum chi-squared method. *Literary and Linguistic Computing* 20: 189-204.
- Windram, HF., Shaw, P., Robinson, P., Howe, CJ. (2008) Dante's *Monarchia* as a test case for the use of phylogenetic methods in stemmatis analysis. *Literary and Linguistic Computing* 23: 443-463.